应用概率统计　第二十七卷
第一期　2011年2月

Chinese Journal of Applied Probability
and Statistics Vol.27 No.1 Feb. 2011

穷举法寻找正交平衡区组设计 *

罗　纯
(上海应用技术学院理学院, 上海, 200235)

潘长缘
(上海君博投资管理有限公司数量分析部, 上海, 200122)

摘　要

对试验次数较小的情形, 我们可以运用穷举法来寻找其可能存在的正交平衡区组设计. 这样做一方面可以尽可能多地找到正交平衡区组设计, 另一方面也可以为今后我们构造试验次数较大的正交平衡区组设计奠定基础. 本文最后给出了试验次数$n=9$以内的部分正交平衡区组设计.

关键词:　正交表, 穷举法, 区组设计, 正交平衡区组设计, 平衡区组正交表.

学科分类号:　O212.6.

§1.　引　　言

在试验设计中, 正交平衡区组设计(等价于广义正交表)[1][2][3][7][8][10]是一种类似于正交拉丁方(或者正交表)的新设计. 根据文[4][5][6]中阐述的利用投影矩阵正交分解技术研究一般设计问题的思想, 发现正交平衡区组设计具有类似正交表的优点, 不但可以大幅度地减少试验次数, 而且保持了数据分析中的正交性质[1][7][8]. 本文只讨论正交平衡区组设计的构造问题. 一组正交的平衡区组设计对应一种等价形式的广义正交表, 而广义正交表可以用来安排实验方案. 如何给出一种合适的方法, 找到正交平衡区组设计组, 构造出相应广义正交表供人们使用, 是试验设计中的一个新问题. 本文对试验次数较小的情形, 尝试运用穷举法来寻找正交平衡区组设计. 当然, 由于穷举法花费时间较长, 对试验次数较大的情形几乎是不可行的.

§2.　正交平衡区组设计的定义

定义 2.1　任意具有v个水平b个区组$B_1=(b_{11},\cdots,b_{1k_1})^T,\cdots,B_b=(b_{b1},\cdots,b_{bk_b})^T$的设计都称为区组设计, 这里$b_{ij}$为整数满足: $1\le b_{ij}\le v$, $j=1,\cdots,k_i$, 而k_i为区组B_i的大小, 即向量B_i的维数. 记这样的区组设计为

$$D_{b\times k}(v)=[B_1,\cdots,B_b]^T=(b_{ij}), \tag{2.1}$$

这里$\boldsymbol{k}=(k_1,\cdots,k_b)^T$为向量. 为了以后叙述方便, 把区组设计$D_{b\times \boldsymbol{k}}(v)=[B_1,\cdots,B_b]^T$视为矩阵(这是一种各行列数可以不相等的多边矩阵[4]), 并把$b\times\boldsymbol{k}$也称为区组设计$D_{b\times\boldsymbol{k}}(v)$的

*教育部高等学校博士学科点专项基金(44K55050)资助.
本文2006年12月30日收到, 2009年6月15日收到修改稿.

阶, 而第i个区组B_i^T对应于矩阵$D_{b\times \boldsymbol{k}}(v)$的第$i$行. 记$n=\sum\limits_{i=1}^{k_i} k_i$, 称为区组设计$D_{b\times \boldsymbol{k}}(v)=[B_1,\cdots,B_b]^T$的试验次数.

对于任意的区组设计$D_{b\times \boldsymbol{k}}(v)=[B_1,\cdots,B_b]^T$, 可以记$I_{B_i}(x)$为区组$B_i$的示性函数, 即

$$I_{B_i}(x)=\begin{cases}1, & \text{如果水平}x\text{在第}i\text{个区组}B_i\text{中出现;}\\ 0, & \text{如果水平}x\text{不在第}i\text{个区组}B_i\text{中出现,}\end{cases}$$

用这个示性函数, 我们可以定义区组设计$D_{b\times \boldsymbol{k}}(v)$的许多性质.

定义 2.2 对于任意一个水平x, $x\in\{1,\cdots,v\}$, 记在区组设计$D_{b\times \boldsymbol{k}}(v)$的第$i$行出现的次数为$r_i(x)$. 若$r_i(x)=0$, 说明水平$x$没有在区组设计$D_{b\times \boldsymbol{k}}(v)$的第$i$行出现. 若$r_i(x)=1$, 说明水平$x$在区组设计$D_{b\times \boldsymbol{k}}(v)$的第$i$行出现了, 但没有重复. 若$r_i(x)>1$, 说明水平$x$在区组设计$D_{b\times \boldsymbol{k}}(v)$的第$i$行出现了, 并且有重复. 而记$r_*(x)=\sum\limits_{i=1}^{b} r_i(x)$为水平$x$在整个区组设计中的重复次数, 简称水平$x$的重复数.

如果两个不同的水平x,y在同一个B_i区组出现, 那么定义它们相遇, 在这个区组的相遇次数定义为$r_i(x)r_i(y)$, 而$\lambda_*(x,y)=\sum\limits_{i=1}^{b} r_i(x)r_i(y)$称为水平$x,y$在整个区组设计中的相遇次数.

将v个水平安排到b个区组$B_1,\cdots,B_b$的一个区组设计$D_{b\times \boldsymbol{k}}(v)$称为平衡的, 假若该设计满足如下二个条件:

1. 对任意水平组合(x,y), $x\neq y$, $x,y\in\{1,\cdots,v\}$, 如下数值是个常数

$$\lambda(x,y)=\sum_{i=1}^{b}\frac{r_i(x)r_i(y)}{k_i}=\lambda>0,\qquad \forall x\neq y,\ x,y\in\{1,\cdots,v\},$$

其中λ称为相遇度, 相遇度必须大于0才理解为平衡. 这个条件称为相遇平衡条件.

2. 对任意水平x; $x\in\{1,\cdots,v\}$, 如下数值是个常数

$$r(x)=\sum_{i=1}^{b}\frac{r_i(x)}{k_i}=r>0,\qquad \forall x\in\{1,\cdots,v\},$$

其中r称为重复度. 这个条件称为重复平衡条件.

记$\boldsymbol{k}=(k_1,\cdots,k_b)^T$, 那么一个平衡的区组设计$D_{b\times \boldsymbol{k}}(v)$都可以用参数$(v,b,r,\boldsymbol{k},\lambda)$刻画.

定义 2.3 一个区组设计称为完全的, 如果在每一个区组内包含设计的所有水平. 一个区组设计称为不完全的, 如果至少存在一个区组在其内不包含设计的所有水平.

定义 2.4 一个区组设计称为组内平衡的, 如果在每一个区组内的元素出现的次数相同.

当$r_i(x)\le 1$, $k_i=k_1<v$, $\forall i\in\{1,\cdots,b\}$时, 对应的平衡不完全的区组设计就是通常的BIB设计.

当$r_i(x) = c,\ k_i = vc,\ \forall i \in \{1, \cdots, b\}$时, 对应的平衡完全的区组设计就是一种广义拉丁矩阵设计[4][9].

当$r_i(x) \le 1,\ k_i < v,\ \forall i \in \{1, \cdots, b\},\ \exists k_i \ne k_{i'}$时, 对应的平衡不完全的区组设计就是PBIB设计.

定义 2.5　两个区组设计称为同阶的, 如果它们各个同样位置(同一行)的区组的大小是相等的.

考虑两个同阶的区组设计$D^1_{b\times \boldsymbol{k}}(v_1) = (b^1_{ij}),\ D^2_{b\times \boldsymbol{k}}(v_2) = (b^2_{ij})$, 对水平组合

$$(x, y), \qquad x \in \{1, \cdots, v_1\},\ y \in \{1, \cdots, v_2\},$$

如果水平组合$(x,y),\ x \in \{1, \cdots, v_1\},\ y \in \{1, \cdots, v_2\}$分别在区组设计$D^1_{b\times \boldsymbol{k}}(v_1) = (b^1_{ij})$, $D^2_{b\times \boldsymbol{k}}(v_2) = (b^2_{ij})$的第$i$区组的相同列出现, 称它们在该区组重复. 水平组合(x,y)在两个区组设计的第i区组的重复次数为集合$\{j; (b^1_{ij}, b^2_{ij}) = (x,y), j = 1, \cdots, k_i\}$内元素的个数, 记为$r^{12}_i(x,y),\ i = 1, \cdots, b$. 而记$r^{12}_*(x,y) = \sum\limits_{i=1}^{b} r^{12}_i(x,y)$, 称为水平组合$(x,y)$ 的在区组设计$D^1_{b\times \boldsymbol{k}}(v_1) = (b^1_{ij}),\ D^2_{b\times \boldsymbol{k}}(v_2) = (b^2_{ij})$中的重复数. 另外定义

$$\lambda^{12}(x,y) = \sum_{i=1}^{b} \frac{r^1_i(x) r^2_i(y)}{k_i},$$

称为水平组合(x,y)在区组设计$D^1_{b\times \boldsymbol{k}}(v_1) = (b^1_{ij}),\ D^2_{b\times \boldsymbol{k}}(v_2) = (b^2_{ij})$中的相遇度.

两个同阶的区组设计$D^1_{b\times \boldsymbol{k}}(v_1) = (b^1_{ij}),\ D^2_{b\times \boldsymbol{k}}(v_2) = (b^2_{ij})$称为正交的, 是说对所有的水平组合$(x,y)$, 它们的重复数与相遇度相等, 即

$$r^{12}_*(x,y) = \lambda^{12}(x,y), \qquad \forall\, x, y. \tag{2.2}$$

易知正交表对应的区组设计满足这个正交性, 并且当$b = 1$时, 其就是平衡设计的定义, 当$b = 1$并且各个设计是组内平衡时, 其就是正交表的定义.

定义 2.6　如果m个平衡的区组设计$D^1_{b\times \boldsymbol{k}}(v_1), \cdots, D^m_{b\times \boldsymbol{k}}(v_m)$是两两正交的, 不妨假设前$r$个区组设计$D^1_{b\times \boldsymbol{k}}(v_1), \cdots, D^l_{b\times \boldsymbol{k}}(v_l)$是完全的, 并且组内平衡, 而后$m-l$个区组设计$D^{l+1}_{b\times \boldsymbol{k}}(v_{l+1}), \cdots, D^m_{b\times \boldsymbol{k}}(v_m)$是不完全的或者组内不平衡的, 记

$$D^t_{b\times \boldsymbol{k}}(v_t) = (B^t_1, \cdots, B^t_b)^T, \qquad t = 1, \cdots, m.$$

定义它们的按行拉长为

$$\mathrm{Vec}(D^t_{b\times \boldsymbol{k}}(v_t)) = ((B^t_1)^T, \cdots, (B^t_b)^T)^T, \qquad t = 1, \cdots, m.$$

那么$n \times (m+1)$矩阵

$$[(1 1^T_{k_1}, \cdots, b 1^T_{k_b})^T, \mathrm{Vec}(D^1_{b\times \boldsymbol{k}}(v_1)), \cdots, \mathrm{Vec}(D^m_{b\times \boldsymbol{k}}(v_m))] \tag{2.3}$$

称为广义正交表, 记为$\mathrm{GL}_n(b^1 v_1 \cdots v_l; v_{l+1} \cdots v_m)$. 并且在水平数相同时, 可以用幂指数$v_1^x$代替$v_1 \overset{x}{\cdots} v_1$.

当$n-1 = (b-1)+(v_1-1)+\cdots+(v_m-1)$时, 称广义正交表$\mathrm{GL}_n(b^1 v_1 \cdots v_l; v_{l+1} \cdots v_m)$是饱和的, 否则称为不饱和的.

上述定义可参看文献[1][2][3][7][8][10].

§3. 算法实现

对试验次数较小的情形, 运用穷举法可以找出所有的正交平衡区组设计. 由于广义正交表可以由一组正交的平衡区组设计改写得到, 因此我们只需要对每一个自由度分解式及其对应的每一个区组分解式, 找出一组正交的平衡区组设计即可. 所谓穷举法, 是指对所有的自由度分解式及其所有的区组分解式, 穷举给出每一个区组设计的每一个元素.

下面两个等式经常用到:

n的自由度分解: 即满足等式$(n-1)=(b-1)+(v_1-1)+\cdots+(v_m-1)$的一组正整数$b,v_1,\cdots,v_m$.

n的区组分解: 把n分解成b(自由度分解式中的b)个正整数的和, 即满足等式$n=k_1+k_2+\cdots+k_b$的一组数$k_1,\cdots,k_b$. 这b个正整数依次表示区组设计的b个区组的元素个数. 给定n的区组分解式, 也就唯一确定了区组设计的形状.

一般地, 给定试验次数n, 要用穷举法寻找正交平衡区组设计, 可以把问题转化为如下六个子问题:

(1) 刻画给出的区组设计;

(2) 检验一个区组设计是否相遇平衡;

(3) 检验一个区组设计是否重复平衡;

(4) 检验给出的两个同阶的区组设计是否相互正交;

(5) 对给定的n的一个自由度分解式, 用穷举的方式找出给定区组分解式下的一组正交平衡区组设计(如果存在的话);

(6) 对给定的n, 穷举所有的自由度分解式及其对应的所有区组分解式, 并按上一问题的方式寻找正交的平衡区组设计组.

逐步解决如上六个子问题, 即解决了整个问题.

为了处理上的方便, 我们将区组设计改写成矩阵. 做法是: 把区组设计中的对应元素作为矩阵的元素, 其中为空的格子用零填充. 这样, 一个区组设计可以对应唯一一个矩阵; 同样地, 可以对应区组设计的矩阵也必然对应唯一一个区组设计. 举例如下:

例 1 $D_{4\times(2,3,4,2)^T}(3)=\begin{pmatrix} 1 & 2 & & \\ 2 & 3 & 1 & \\ 1 & 2 & 2 & 1 \\ 3 & 1 & & \end{pmatrix}$是一个区组设计, 其对应的矩阵为

$$A=\begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 3 & 1 & 0 \\ 1 & 2 & 2 & 1 \\ 3 & 1 & 0 & 0 \end{pmatrix}.$$

下面我们将一一解决上述问题.

(1) 刻画给出的区组设计

我们通常关心区组设计的五个参数. 要计算这五个参数, 我们首先要对该区组设计进行简单的刻画. 一般地, 考虑用矩阵或者向量的形式存储这些信息.

先把给定区组设计转化成矩阵A.

我们将矩阵A的b个不同行的非零元素分别记录下来, 用一个与矩阵A相同阶数的新矩阵C_A存储(其中矩阵C_A的行与矩阵A的行对应, 并且按矩阵A的行中非零元素出现的顺序依次记录, 其余元素设为零). 与此同时, 记录矩阵A各行不同元素的个数, 得到一个b维列向量s_A, 列向量s_A的行也对应矩阵A的行. 对矩阵A每行的每个非零元素, 记录其在同一行中出现的次数, 也用一个与矩阵A相同阶数的新矩阵D_A存储下来, 矩阵D_A中的元素分别记录矩阵C_A的对应位置上的非零元素在矩阵A中对应行中重复出现的次数, 若矩阵C_A的对应位置上的元素为零, 则矩阵D_A中该元素也为零.

一般地, 单个区组设计的基本信息可以由上述矩阵或向量C_A, s_A, D_A完全描述; 在讨论两个区组设计是否正交时, 则仍然需要考虑矩阵A. 举例如下:

例 2　给出一个区组设计

$$D_{4\times(2,3,4,4)^T}(3) = \begin{pmatrix} 1 & 2 & & \\ 2 & 3 & 1 & \\ 1 & 2 & 2 & 1 \\ 3 & 1 & 3 & 2 \end{pmatrix},$$

其对应矩阵A:

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 3 & 1 & 0 \\ 1 & 2 & 2 & 1 \\ 3 & 1 & 3 & 2 \end{pmatrix}.$$

依据该算法, 可得:

$$C_A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 3 & 1 & 0 \\ 1 & 2 & 0 & 0 \\ 3 & 1 & 2 & 0 \end{pmatrix}; \qquad s_A = \begin{pmatrix} 2 \\ 3 \\ 2 \\ 3 \end{pmatrix}; \qquad D_A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 2 & 2 & 0 & 0 \\ 2 & 1 & 1 & 0 \end{pmatrix}.$$

(2) 判断一个区组设计是否相遇平衡

相遇平衡指的是任意水平组合(x,y), $x \neq y$, $x,y \in \{1,\cdots,v\}$,

$$\lambda(x,y) = \sum_{i=1}^{b} \frac{r_i(x)r_i(y)}{k_i} = \lambda > 0, \qquad \forall\, x \neq y,\ x,y \in \{1,\cdots,v\},$$

其中λ就是相遇度, 相遇度必须大于0才理解为平衡.

仍然将给定的区组设计处理成矩阵A. 此处只需要记录矩阵A中的元素组合(x,y), $x\neq y$ (无序组合)在整个矩阵各行中的$r_i(x)r_i(y)/k_i$之和, 并判断其是否相等.

伴随着第一个问题的解决, 我们已掌握了区组设计所对应的矩阵A一些基本信息, 例如C_A, s_A, D_A.

由于矩阵C_A记录了矩阵A中每行出现的非零元素, 那么对矩阵C_A (矩阵A中各行出现的非零元素在矩阵C_A的对应行都出现一次)每行中出现的所有元素的组合都计算$r_i(x)r_i(y)/k_i$ (无序). 将所有的组合(x,y)在整个矩阵中出现的行中计算得到的$r_i(x)r_i(y)/k_i$累加到矩阵G_A的元素$g(x,y)$及$g(y,x)$, 这样得到一个v阶主对角线元素为零(其余元素也可能为零)的对称矩阵G_A (此处v代表矩阵对应区组设计的水平数). 具体处理时可以考虑如下累加运算: 首先让矩阵G_A为v阶零矩阵; 然后对矩阵C_A的b个不同行分别考察, 出现一次组合(x,y)则让矩阵G_A的元素$g(x,y)$及$g(y,x)$都在原来的基础上增加$r_i(x)r_i(y)/k_i$. 此处不对矩阵A而对矩阵C_A做处理的好处在于矩阵C_A中的各行元素不重复.

判断矩阵G_A中除主对角线元素以外的其他元素是否相等(在具体处理上, 只要判断整个矩阵G_A的除主对角元素之外的其他元素中的最大值与最小值是否相等即可). 如果相等, 则表明矩阵A对应的区组设计水平间平衡; 如果不相等, 则说明矩阵A对应的区组设计不是水平间平衡的. 举例如下:

例 3 给出一个区组设计

$$D_{4\times(3,4,4,4)^T}(3)=\begin{pmatrix}1&2&3&\\1&2&2&1\\3&2&3&2\\3&1&3&1\end{pmatrix},$$

其对应的矩阵为A:

$$A=\begin{pmatrix}1&2&3&0\\1&2&2&1\\3&2&3&2\\3&1&3&1\end{pmatrix}.$$

依据该算法, 可得

$$C_A=\begin{pmatrix}1&2&3&0\\1&2&0&0\\3&2&0&0\\3&1&0&0\end{pmatrix};\qquad s_A=\begin{pmatrix}3\\2\\2\\2\end{pmatrix};$$

$$D_A=\begin{pmatrix}1&1&1&0\\2&2&0&0\\2&2&0&0\\2&2&0&0\end{pmatrix};\qquad G_A=\begin{pmatrix}0&4/3&4/3\\4/3&0&4/3\\4/3&4/3&0\end{pmatrix}.$$

矩阵G_A除主对角元素外的其余元素都为$4/3$, 由此可知矩阵A对应的区组设计是相遇平衡的.

(3) 判断一个区组设计是否重复平衡

重复平衡是指对任意水平x; $x \in \{1, \cdots, v\}$,

$$r(x)=\sum_{i=1}^{b}\frac{r_i(x)}{k_i}=r>0, \qquad \forall\, x\in\{1,\cdots,v\},$$

其中r就是重复度.

把给定的区组设计处理成矩阵A. 我们只需要对每一个元素x, 求出其在该矩阵对应的指标:

$$r(x)=\sum_{i=1}^{b}\frac{r_i(x)}{k_i}, \qquad \forall\, x\in\{1,\cdots,v\},$$

并判断是否每个$r(x)$相等, $(x=1,\cdots,v)$. 具体处理时, 只要判断$r(x)$ $(x=1,\cdots,v)$中最大值与最小值是否相等. 如果相等, 则说明矩阵对应的区组设计重复平衡.

前面已经得到了矩阵C_A、s_A和D_A. 分别对矩阵C_A中元素x出现的每一行, 求出

$$\omega_i(x)=\frac{r_i(x)}{k_i}, \qquad i\in\{1,\cdots,b\},\ x\in\{1,\cdots,v\}.$$

然后对每个水平按行相加, 即得最终的r_x, $(x=1,\cdots,v)$. 记$r_A=(r_1,\cdots,r_v)^T$为矩阵A所对应的区组设计中各水平重复度向量. 举例如下:

例 4 给出一个区组设计

$$D_{4\times(3,4,4,4)^T}(3)=\begin{pmatrix} 3 & 1 & 2 & \\ 1 & 1 & 3 & 3 \\ 1 & 2 & 2 & 1 \\ 3 & 2 & 3 & 2 \end{pmatrix},$$

其对应的矩阵为A:

$$A=\begin{pmatrix} 3 & 1 & 2 & 0 \\ 1 & 1 & 3 & 3 \\ 1 & 2 & 2 & 1 \\ 3 & 2 & 3 & 2 \end{pmatrix}.$$

依据该算法, 可得:

$$C_A=\begin{pmatrix} 3 & 1 & 2 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 3 & 2 & 0 & 0 \end{pmatrix}; \qquad s_A=\begin{pmatrix} 3 \\ 2 \\ 2 \\ 2 \end{pmatrix}; \qquad D_A=\begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \end{pmatrix};$$

$$G_A=\begin{pmatrix}0&4/3&4/3\\4/3&0&4/3\\4/3&4/3&0\end{pmatrix};\qquad r_A=\begin{pmatrix}4/3\\4/3\\4/3\end{pmatrix}.$$

因此, 由G_A可知, 矩阵A对应的区组设计满足相遇平衡条件; 由向量r_A可知, 矩阵A对应的区组设计满足重复平衡条件.

(4) 检验给出的两个同阶的区组设计是否相互正交

所谓相互正交, 是指给出的两个同阶区组设计$D^1_{b\times\boldsymbol{k}}(v_1)$和$D^2_{b\times\boldsymbol{k}}(v_2)$的相遇度与重复数相等, 即:

$$\begin{aligned}\lambda^{12}(x,y)&=\sum_{i=1}^{b}\frac{r_i^1(x)r_i^2(y)}{k_i}=\sum_{i=1}^{b}\sum_{j=1}^{k_i}I_{b^1_{ij}}(x)I_{b^2_{ij}}(y)\\&=r_*^{12}(x,y),\qquad\forall\,x\in D^1_{b\times\boldsymbol{k}}(v_1),\ y\in D^2_{b\times\boldsymbol{k}}(v_2).\end{aligned}\tag{3.1}$$

我们仍然把区组设计看成矩阵A和B.

这样, 先对A和B分别进行刻画可以得到关于这两个矩阵的一些基本信息, 例如C_A, s_A, D_A, C_B, s_B, D_B.

我们对$\forall\,x\in A$, $x\neq 0$, $y\in B$, $y\neq 0$, 分别求出$\lambda_{12}(x,y)$以及$r_*^{12}(x,y)$, 并判断它们是否相等即可. 如果相等, 则说明矩阵A和B所对应的区组设计相互正交.

矩阵A和B中每行相同位置元素组合(x,y)重复出现的次数为$r_i^{12}(x,y)$; 对每一对元素组合(x,y), $\forall\,x\in A$, $y\in B$分行求出

$$t_i^{12}(x,y)=\frac{r_i^1(x)r_i^2(y)}{k_i}.\tag{3.2}$$

然后我们对元素组合(x,y)按行进行累加, 可以得到两个$v_1\times v_2$阶的矩阵R_{AB}, T_{AB}, 其第(x,y)个元素表示在矩阵A与矩阵B对应区组设计中的水平x与y的重复数与相遇度.

因此, 只需判断矩阵R_{AB}和矩阵T_{AB}是否相等即可. 如果矩阵R_{AB}和矩阵T_{AB}相等, 则表明矩阵A和B对应的区组设计正交; 反之, 则不正交. 举例如下:

例 5 给出两个区组设计

$$D^1_{3\times(2,2,2)^T}(2)=\begin{pmatrix}1&2\\1&2\\1&2\end{pmatrix},\qquad D^2_{3\times(2,2,2)^T}(3)=\begin{pmatrix}1&2\\2&3\\3&1\end{pmatrix}.$$

其各自对应的矩阵如下:

$$A=\begin{pmatrix}1&2\\1&2\\1&2\end{pmatrix},\qquad B=\begin{pmatrix}1&2\\2&3\\3&1\end{pmatrix}.$$

由于矩阵A和B对应的平衡区组设计的最大水平分别为2和3, 因此得到如下两个矩阵:

$$R_{AB} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad T_{AB} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

矩阵R_{AB}与矩阵T_{AB}相等. 因此, 矩阵A和B对应的区组设计相互正交.

(5) 对给定的n的一个自由度分解式, 穷举法找出给定区组分解式下的一组正交平衡区组设计

我们知道, 为了得到一组正交平衡区组设计, 可以按照自由度分解式的顺序寻找. 对给定的自由度分解式和给定区组分解式唯一确定的区组设计形状, 尝试找出一组正交平衡区组设计.

对给定的试验次数n的一个自由度分解式$n-1=(b-1)+(v_1-1)+(v_2-1)+\cdots+(v_m-1)$. b表明每个区组设计有b个区组, $v_1, v_2, \cdots, v_m$对应我们要寻找的一组正交平衡区组设计依次对应的水平数. 由于试验次数为n, 我们对n作区组分解, 把n分解成b个正整数的和, 这b个正整数代表我们寻找的正交平衡区组设计的b个区组中各区组元素的个数. 这样, 区组设计(矩阵)的形状就确定了. 现在我们的目标是找出m个相互正交的平衡区组设计(矩阵). 由自由度分解式我们知道要寻找的平衡区组设计对应的水平数依次为$v_1, v_2, \cdots, v_m$, 这表明我们要寻找的一组矩阵的最大元素依次为$v_1, v_2, \cdots, v_m$.

我们用穷举的方式分别对这m个矩阵找出其对应区组设计满足平衡条件的所有矩阵, 然后对所有可能的矩阵组合情形, 判断其对应的平衡区组设计是不是两两相互正交. 如果发现一组矩阵其对应平衡区组设计相互正交, 则给出结果即可.

当然, 我们不能保证对每种确定的矩阵形状都能找出m个矩阵, 使得它们对应的区组设计是平衡且相互正交的, 甚至我们不能保证对n的每个自由度分解式能够找出一组结果. 这可能有多方面的原因: 譬如对某种区组形状可能根本不存在某一水平的平衡区组设计, 甚至可能对某些试验次数n根本不存在平衡的区组设计.

根据自由度分解式找出来的一组正交平衡区组设计对应广义正交表中最大的列数(饱和). 当然对于某些自由度分解式及其某个区组分解式, 不能找到一组饱和的正交平衡区组设计, 但是可能可以找到比饱和情形数目(m)少一些的一组正交平衡区组设计, 它们同样可以用来改写成广义正交表, 也可以在实际试验中使用. 因此, 我们对这种情形也应该加以考虑. 这样, 我们最后给出的广义正交表不一定都是饱和的, 但由于穷举了所有情形, 它的列一定是最大数目的.

用穷举法产生第t个矩阵时, 需要判断$1, \cdots, v_t$是否在该矩阵中全部出现($t=1, \cdots, m$). 我们只需要检测出被产生的矩阵的最大元素为v_t, 那么这个矩阵一定包括了从1到v_t之间的任何一个数(这正是我们要寻找的区组设计所要求的), 这是因为判断矩阵所对应的区组设计是否平衡的程序保证了这一点.

(6) 对给定的n, 尝试找出所有自由度分解式下的所有区组分解式的一组正交平衡区组设计

对于一个给定的试验次数n, 我们可以穷举出它的所有自由度分解式, 这一点用计算机不难办到. 当然, 一旦给定了自由度分解式, 我们又可以根据其中的b给出n的所有区组分解

式, 这个分解完全是为了确定区组设计的形状. 这个过程, 也不难用计算机实现. 接下来, 只需要对n的每一个自由度分解式及其相应的每一确定区组分解式, 寻找一组正交的平衡区组设计即可.

我们运用穷举法给出n的所有自由度分解式, 并穷举出n相应的所有区组分解式, 然后对每一种情形运用解决上一问题的方法寻找结果.

当然我们根据自由度分解式中的b, 对n做区组分解时可以只给出n的那些单调不减分解. 这样做一方面可以减少寻找时间; 另一方面n的其他区组分解仅仅对应了交换区组设计的某些区组顺序, 而这并不影响结果的存在性.

事实上, 如果对n的某个单调不减区组分解式得到了一组正交平衡区组设计, 那么我们对这一组正交平衡区组设计做区组顺序变换, 这样就得到一组新的区组设计, 这组新的区组设计仍然是平衡且正交的. 如果对这组正交平衡区组设计做所有可能的区组顺序变换, 就可以得到许多新的正交平衡区组设计组. 反过来, 对n的任意某个区组分解式下得到的一组正交平衡区组设计, 我们都可以通过交换区组顺序的方法使区组分解式变成单调不减的, 并且这组区组设计仍然是平衡且正交的. 因此, 在这里, 我们只需穷举n的那些单调不减区组分解即可.

§4. 主要结果

对给定的某个自由度分解式以及某个确定的区组分解式, 如果找到一组矩阵, 使得它们对应的区组设计平衡且正交, 那么, 我们就可以将它们改写成广义正交表. 需要注意的是, 由一组正交平衡区组设计改写成的广义正交表的列数比这组正交平衡区组设计的个数多一列, 这是因为广义正交表的第一列完全可以根据区组分解式写出. 在这里, 我们对同一自由度分解式以及同一区组分解式的结果只给出一张广义正交表. 事实上同一自由度分解式以及同一区组分解式下可能存在多组结果, 例如对一组给出的结果通过一些区组间交换、正交平衡区组设计顺序交换以及组内数码置换就可以得到多组不同的结果. 这里我们将可以通过区组间交换、正交平衡区组设计顺序交换以及组内数码置换就可以得到多组不同的结果归为一类, 这样就只对每一类给出一张广义正交表. 至于同一自由度分解式以及同一区组分解式下是否存在多类, 将在以后讨论.

通过计算机穷举, 我们给出以下$n=9$以内的广义正交表:

试验次数$n=4,5,6$的广义正交表

$GL_4(3^1;2^1)$	$GL_4(2^3)$	$GL_5(3^1;2^2)$	$GL_6(4^1;3^1)$	$GL_6(5^1;2^1)$	$GL_6(3^1;2^1)$
$\begin{pmatrix}1&1\\2&2\\3&1\\3&2\end{pmatrix}$	$\begin{pmatrix}1&1&1\\1&2&2\\2&1&2\\2&2&1\end{pmatrix}$	$\begin{pmatrix}1&1&2\\2&2&1\\2&2&2\\3&2&1\\3&1&1\end{pmatrix}$	$\begin{pmatrix}1&1\\2&2\\3&3\\4&1\\4&2\\4&3\end{pmatrix}$	$\begin{pmatrix}1&1\\2&1\\3&2\\4&2\\5&1\\5&2\end{pmatrix}$	$\begin{pmatrix}1&2\\2&1\\2&2\\3&1\\3&1\\3&1\end{pmatrix}$

试验次数$n=6$的广义正交表

$GL_6(2^2)$	$GL_6(3^1 2^1;3^1)$	$GL_6(4^1;2^2)$	$GL_6(2^1 3^1;2^1)$	$GL_6(3^1;2^3)$	$GL_6(3^1;2^3)$
$\begin{pmatrix}1&1\\1&2\\2&1\\2&2\\2&2\\2&1\end{pmatrix}$	$\begin{pmatrix}1&1&1\\1&2&2\\2&1&2\\2&2&3\\3&1&3\\3&2&1\end{pmatrix}$	$\begin{pmatrix}1&1&1\\2&2&2\\3&1&1\\3&2&2\\4&1&2\\4&2&1\end{pmatrix}$	$\begin{pmatrix}1&2&1\\1&3&1\\1&1&2\\2&2&2\\2&3&2\\2&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1\\2&2&2&2\\3&1&1&1\\3&1&2&2\\3&2&1&2\\3&2&2&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1\\1&1&1&2\\2&1&2&1\\2&2&2&1\\3&2&1&2\\3&2&2&2\end{pmatrix}$

试验次数$n=7$的广义正交表

$GL_7(4^1;3^1)$	$GL_7(5^1;2^2)$	$GL_7(3^1;2^2)$	$GL_7(4^1;2^2)$	$GL_7(3^1;2^4)$
$\begin{pmatrix}1&1\\2&2\\3&3\\3&3\\4&2\\4&3\\4&1\end{pmatrix}$	$\begin{pmatrix}1&1&1\\2&1&2\\3&2&2\\4&2&1\\4&2&2\\5&2&1\\5&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&2\\1&2&2\\2&2&1\\2&2&2\\3&1&1\\3&1&1\\3&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1\\2&1&2\\2&2&1\\3&2&2\\3&2&2\\4&2&2\\4&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&2\\2&2&2&2&1\\2&2&2&2&2\\3&1&2&2&1\\3&2&1&2&1\\3&2&2&1&1\\3&1&1&1&1\end{pmatrix}$

试验次数$n=8$的广义正交表(一)

$GL_8(5^1;4^1)$	$GL_8(7^1;2^1)$	$GL_8(3^1;2^1)$	$GL_8(3^1;2^1)$	$GL_8(3^1;2^1)$	$GL_8(4^1;2^1)$	$GL_8(4^1;2^1)$	$GL_8(4^1;2^1)$
$\begin{pmatrix}1&1\\2&2\\3&3\\4&4\\5&1\\5&2\\5&3\\5&4\end{pmatrix}$	$\begin{pmatrix}1&1\\2&1\\3&1\\4&2\\5&2\\6&2\\7&1\\7&2\end{pmatrix}$	$\begin{pmatrix}1&2\\2&2\\2&1\\3&1\\3&1\\3&1\\3&1\\3&1\end{pmatrix}$	$\begin{pmatrix}1&2\\2&1\\3&2\\3&2\\3&2\\3&1\\3&1\\3&1\end{pmatrix}$	$\begin{pmatrix}1&2\\1&1\\2&2\\2&2\\2&2\\3&1\\3&1\\3&1\end{pmatrix}$	$\begin{pmatrix}1&2\\2&1\\3&2\\3&1\\4&2\\4&2\\4&1\\4&1\end{pmatrix}$	$\begin{pmatrix}1&2\\2&2\\2&1\\3&2\\3&1\\4&1\\4&1\\4&1\end{pmatrix}$	$\begin{pmatrix}1&2\\1&2\\2&2\\2&1\\3&2\\3&1\\4&1\\4&1\end{pmatrix}$

试验次数$n=8$的广义正交表(二)

$GL_8(5^1;2^1)$	$GL_8(5^1;2^1)$	$GL_8(4^1;3^2)$	$GL_8(6^1;2^2)$	$GL_8(4^1;2^2 3^1)$	$GL_8(5^1;2^3)$
$\begin{pmatrix}1&2\\2&2\\3&1\\4&1\\5&2\\5&2\\5&1\\5&1\end{pmatrix}$	$\begin{pmatrix}1&2\\2&2\\3&1\\4&2\\4&1\\5&1\\5&1\\5&1\end{pmatrix}$	$\begin{pmatrix}1&3&3\\2&2&2\\3&3&1\\3&2&1\\3&1&1\\4&1&3\\4&1&2\\4&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1\\2&1&1\\3&2&2\\4&2&2\\5&1&1\\5&2&2\\6&1&2\\6&2&1\end{pmatrix}$	$\begin{pmatrix}1&2&2&3\\2&2&1&2\\2&1&2&2\\3&2&2&1\\3&1&1&1\\4&1&1&3\\4&1&1&2\\4&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1\\2&1&1&1\\3&2&2&2\\4&2&2&2\\5&1&1&1\\5&1&2&2\\5&2&1&2\\5&2&2&1\end{pmatrix}$

试验次数$n=8$的广义正交表(三)

$GL_8(5^1;2^3)$	$GL_8(2^3;3^1)$	$GL_8(3^1;2^3)$	$GL_8(2^14^12^3)$
$\begin{pmatrix}1&2&2&2\\2&2&2&2\\3&2&1&1\\3&1&1&1\\4&1&2&1\\4&1&1&1\\5&1&1&2\\5&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&2&1&2\\1&1&2&3\\2&2&2&1\\2&2&2&1\\2&2&1&3\\2&1&2&2\\2&1&1&1\\2&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&2&2&2\\2&1&1&1\\2&1&1&1\\2&1&1&1\\3&2&2&1\\3&2&1&2\\3&1&2&2\\3&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&1\\1&2&1&2&2\\1&3&2&1&2\\1&4&2&2&1\\2&1&2&2&2\\2&2&2&1&1\\2&3&1&2&1\\2&4&1&1&2\end{pmatrix}$

试验次数$n=8$的广义正交表(四)

$GL_8(4^1;2^4)$	$GL_8(2^5)$	$GL_8(3^1;2^5)$	$GL_8(2^7)$
$\begin{pmatrix}1&2&2&2&1\\1&2&2&1&2\\2&2&1&2&2\\2&1&2&2&2\\3&2&2&1&1\\3&1&1&1&1\\4&1&1&2&2\\4&1&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&2&1&1&2\\1&1&2&2&1\\2&2&2&2&2\\2&2&2&1&1\\2&2&1&2&1\\2&1&2&1&2\\2&1&1&2&2\\2&1&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&2&2&2&2&2\\1&2&1&2&2&2\\2&2&2&1&1&1\\2&1&2&1&1&1\\3&1&1&2&2&1\\3&1&1&2&1&2\\3&1&1&1&2&2\\3&1&1&1&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&1&1&1&1&1\\1&1&1&2&2&2&2\\1&2&2&1&1&2&2\\1&2&2&2&2&1&1\\2&1&2&1&2&1&2\\2&1&2&2&1&2&1\\2&2&1&1&2&2&1\\2&2&1&2&1&1&2\end{pmatrix}$

试验次数$n=9$的广义正交表

$GL_9(7^1;3^1)$	$GL_9(2^1;3^1)$	$GL_9(4^1;3^1)$	$GL_9(5^1;3^1)$	$GL_9(6^1;3^1)$	$GL_9(4^1;3^2)$	$GL_9(3^4)$
$\begin{pmatrix}1&1\\2&1\\3&2\\4&2\\5&3\\6&3\\7&2\\7&3\\7&1\end{pmatrix}$	$\begin{pmatrix}1&1\\1&2\\1&3\\2&1\\2&2\\2&2\\2&3\\2&3\\2&1\end{pmatrix}$	$\begin{pmatrix}1&1\\2&2\\3&3\\4&1\\4&2\\4&2\\4&3\\4&3\\4&1\end{pmatrix}$	$\begin{pmatrix}1&1\\2&2\\3&3\\4&1\\4&2\\4&3\\5&2\\5&3\\5&1\end{pmatrix}$	$\begin{pmatrix}1&1\\2&2\\3&3\\4&1\\4&2\\5&2\\5&3\\6&3\\6&1\end{pmatrix}$	$\begin{pmatrix}1&1&2\\1&2&1\\2&1&3\\2&3&1\\3&2&3\\3&3&2\\4&2&2\\4&3&3\\4&1&1\end{pmatrix}$	$\begin{pmatrix}1&1&2&2\\1&2&1&3\\1&3&3&1\\2&1&3&3\\2&2&2&1\\2&3&1&2\\3&2&3&2\\3&3&2&3\\3&1&1&1\end{pmatrix}$

本文给出的广义正交表我们也称之为平衡区组正交表, 以后可以和正交表一样进行数据分析.

参 考 文 献

[1] 潘长缘, 正交平衡区组设计, 华东师范大学硕士学位论文, 2009.

[2] 张应山, 张晓琴, 潘长缘, 田金亭, 茆诗松, 正交平衡区组设计的基本理论, 宁波: 第一届中国"试验设计与质量改进"会议, 2007.

[3] Zhang, Y.S., Wang, X.D., Zhang, X.Q., Pan, C.Y., Chen, X.P., Wang, H., Tian, J.T. and Mao, S.S., Essential theory of orthogonal balanced block designs, 南京: 第六届海峡两岸统计与概率研讨会, 2009.

[4] 张应山, *多边矩阵理论*, 北京: 中国统计出版社, 1993.

[5] 张应山, 正交表的数据分析及其构造, 华东师范大学博士学位论文, 2006.

[6] Zhang, Y.S., Lu, Y.Q. and Pang, S.Q., Orthogonal arrays obtained by orthogonal decomposition of projection matrices, *Statistica Sinica*, **9**(1999), 595–604.

[7] 张应山, 张晓琴, 潘长缘, 田金亭, 茆诗松, 正交平衡区组设计的统计分析, 宁波: 第一届中国"试验设计与质量改进"会议, 2007.

[8] Zhang, Y.S., Zhang, X.Q., Pan, C.Y., Wang, X.D., Chen, X.P., Wang, H., Tian, J.T. and Mao S.S., Statistical analysis of orthogonal balanced block designs, 贵阳: 第四届全国生存分析与应用统计研讨会, 2008.

[9] 张应山, 卢一强, 广义拉丁矩阵的计数, *河南师范大学学报(自然科学版)*, **25(1)**(1997), 9–11.

[10] 田金亭, 张应山, 张晓琴, 潘长缘, 甘媛源, 平衡区组正交表与正交表的比较及应用, *数学的实践与认识*, **39(22)**(2009), 59–67.

Method of Exhaustion to Search Orthogonal Balanced Block Designs

Luo Chun

(*College of Sciences, Shanghai Institute of Technology, Shanghai, 200235*)

Pan Changyuan

(*Department of Quantitative Analysis, Junbo Capital Management CO., LTD, Shanghai, 200122*)

When the run size of experiment is relative small, it is possible to use the method of exhaustion to select its possible existing orthogonal balanced block designs. In this way, we can find out both orthogonal balanced block designs as more as possible and some information on how to construct orthogonal balanced block designs for some experiment of big run size. Some result of experiment run size of no more than 9 is given out.

Keywords: Orthogonal arrays, method of exhaustion, block designs, orthogonal balanced block designs, balanced block orthogonal arrays.

AMS Subject Classification: 62K15, 05B15.